

Author's preprint, published version at: <https://dl.acm.org/citation.cfm?id=3290605.3300771>

Saliency Deficit and Motion Outlier Detection in Animated Scatterplots

Rafael Veras

University of Ontario Institute of Technology
Oshawa, ON, Canada
rafael.verasguimaraes@uoit.ca

Christopher Collins

University of Ontario Institute of Technology
Oshawa, ON, Canada
christopher.collins@uoit.ca

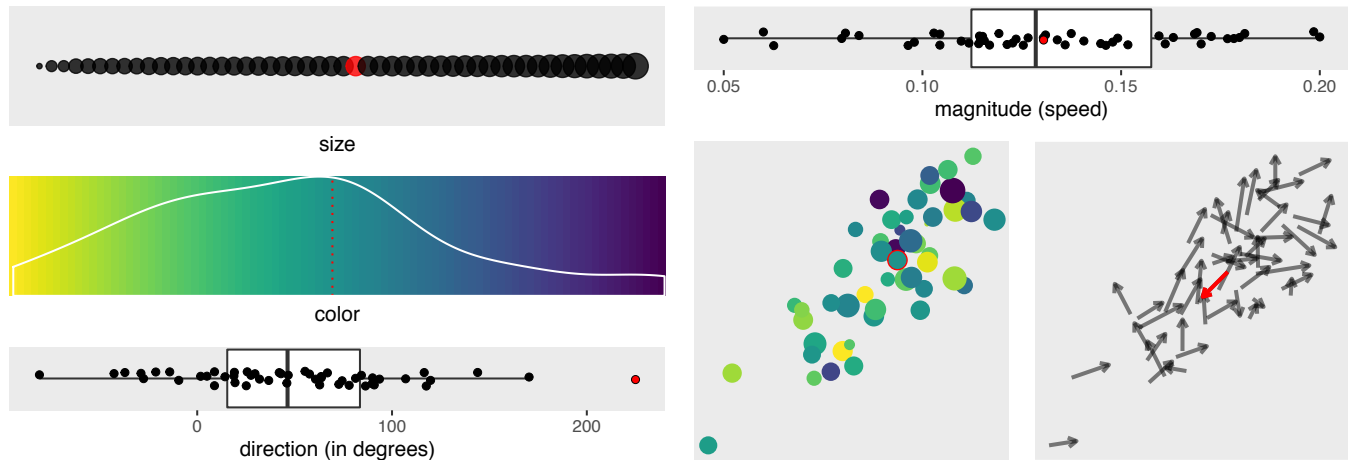


Figure 1: Feature distributions in a typical direction stimuli. Values are sampled from independent distributions. The target point (marked in red) receives an outlying value in the relevant dimension (either direction or speed). Depending on the condition, the target can have mean or maximum (salient) values in the irrelevant dimensions. For instance, in the condition *+color*, the target has salient color. The scatterplot in the bottom left is the first frame; the arrows in the bottom right represent the displacement between the initial and final frames, which is animated.

ABSTRACT

We report the results of a crowdsourced experiment that measured the accuracy of motion outlier detection in multivariate, animated scatterplots. The targets were outliers either in speed or direction of motion, and were presented with varying levels of saliency in dimensions that are irrelevant to the task of motion outlier detection (e.g., color, size, position). We found that participants had trouble finding the outlier when it lacked irrelevant salient features and that

visual channels contribute unevenly to the odds of an outlier being correctly detected. Direction of motion contributes the most to accurate detection of speed outliers, and position contributes the most to accurate detection of direction outliers. We introduce the concept of saliency deficit in which item importance in the data space is not reflected in the visualization due to a lack of saliency. We conclude that motion outlier detection is not well supported in multivariate animated scatterplots.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-5970-2/19/05...\$15.00

<https://doi.org/10.1145/3290605.3300771>

CCS CONCEPTS

• Human-centered computing → Empirical studies in visualization;

ACM Reference Format:

Rafael Veras and Christopher Collins. 2019. Saliency Deficit and Motion Outlier Detection, in *Animated Scatterplots*. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3290605.3300771>

1 INTRODUCTION

In this paper, we investigate questions related to the independence of visual dimensions in animated scatterplots. We often seek to encode data in as many visual variables as possible, and this strategy has been extended to scatterplots with the use of color, size, and motion. Here we question the accuracy of the basic task of motion outlier detection in the complex scenes formed by animated multivariate scatterplots. Does the saliency of non-motion features impact the detection of motion outliers? Can we put motion outliers in a state where they are hard to detect by simply changing their color, size, or position? If so, in visualizations where observing change is a relevant task the variations in data point saliency will hinder or amplify the local perception of change, turning the encoding unreliable.

The perception literature has abundant studies on the performance of search tasks in static and moving scenes [7, 8, 21, 33, 35]. However, psychology studies are difficult to comprehend by non-experts, and their low level makes it difficult to extract implications for visualization design. Nonetheless, these controlled experiments produced general results that support useful rules of thumb; for instance, targets among uniform distractors are much easier to detect than when the distractors have high variance [7]. This rule captures well the results of “pre-attention” experiments with single and conjunction static features (e.g., color), and with motion components (speed and direction). Detection of speed and direction outliers in displays where no other features compete is considered efficient, and the effects of speed on direction and vice-versa are well studied [25]. However, detecting speed and direction targets in scenes where many other channels are used is not well studied.

In the second edition of his book, Ware warned that studies on perceptual independence among three or more visual channels were rare [34]. Almost 15 years later, our understanding of these interactions and their implications to visualization is insufficient, and fewer are the studies that involve motion in visualization. Progress recently has been made in revising rankings of encoding effectiveness [17, 22]. While these have great practical application, they do not seek to explain the fundamental phenomena driving performance results.

Among the powerful concepts that may help us unveil the roots of problems in the visual mapping of data is visual saliency. In this paper, we contribute an experiment aimed at measuring the gap in motion outlier detection accuracy between salient and non-salient outliers. We simulate animated scatterplots that contain either a speed outlier or a direction outlier. Then we vary the number of static features that, in addition to motion, are salient in these outliers.

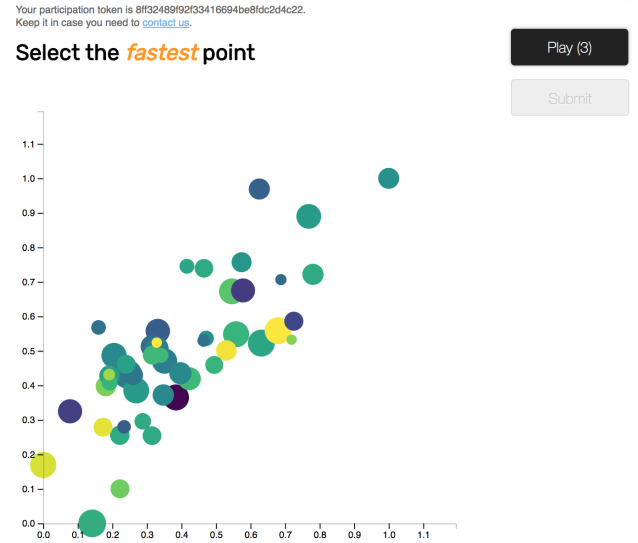


Figure 2: Snapshot of the interface for the speed task. The direction task asked “Select the point that moves in the most deviant direction.”

We find that motion outliers that have additional salient features are much more likely to be correctly identified than non-salient outliers. Our results show that motion is not immune from the interference of other dimensions and suggest that motion outlier detection is unreliable in multivariate animated scatterplots. We proceed to define the notion of *saliency deficit*: a state where the saliency profile in a visualization scene impairs the effectiveness of a visualization task; and suggest that saliency deficit models can help the automatic identification of saliency-boosting opportunities in visualizations.

2 RELATED WORK

In this research, we are interested in the role saliency plays in motion outlier identification. While this question has wide-ranging applications, we constrain our investigation to animated scatterplots. In this section, we will review the related work in perception for information visualization, the use of animated scatterplots, and the recent trend of developing empirical perception models for visualization.

Perception

Visual attention research investigates the limits of attention of the human visual system and has produced a number of theories that explain the mechanisms of visual information processing (see Healey and Enns [13] for a review). Feature integration theory proposes that scenes are initially processed as many separable basic dimensions (e.g., color, motion, orientation), which are later integrated to

form more complex objects [32]. Without focused attention, features remain separated. As a consequence of this mechanism, searches for basic features occur in parallel and are fast, while searches for conjunction features, which involve more than one dimension (e.g., a red circle in a scene with red squares and blue circles), occur serially and thus slow down as the number of objects present in the scene increase. Visual search experiments usually ask participants to determine whether a target is present in a scene with distractors, and the number of distractors is manipulated. Reaction times (RT) and accuracy are recorded, and results are summarized as the slope of the linear relationship between the response and the number of distractors. Parallel searches have slope close to 0. Frequently, the term “popout” is used to describe the easy identification of targets in these searches.

While many experiments corroborate feature integration theory, other experiments found that some conjunction searches are too efficient to be serial searches. For instance, motion-shape targets can be detected in parallel, suggesting the existence of a *motion filtering process*, which effectively subsets the scene, reducing the search task to a simple feature search on moving items [21, 33]. Aiming at explaining these problematic cases, the theory of guided search posits that the goals of the viewer play a large role in visual search, with activation maps (“heatmap” representations of the visual space storing the likelihood of locations containing a target) being constructed with bottom-up and top-down information. Top-down processes are cognitive, driven by users tasks and goals, while bottom-up processes are driven by sensory information. Guided search theory suggests that the difference in performance between single feature and conjunction tasks is due to the amount of guidance that bottom-up processes can provide [35]. Thus top-down guidance is the reason “fast” conjunction searches exist.

The impact of color on motion discrimination is well studied. Both hue and luminance have been shown to independently enable *apparent motion* of simple objects when they are displayed in different positions in successive frames, prompting debate as to whether or not color and motion are processed by separate pathways [23]. Croner and Albright [6] found that hue saliency and luminance saliency aid the discrimination of motion direction; that is, participants detect more accurately targets moving in the same direction among distractors moving in random directions when the targets have distinct hue or luminance, which may suggest that color segmentation of the scene occurs prior to motion discrimination, a process opposite to the motion filtering mentioned above.

The statistical saliency model (SSM) [25] seeks to explain motion popout phenomena with a simple statistical measure that quantifies the saliency of targets with respect to the distractors in the scene. The SSM explains the following

asymmetries in motion popout phenomena: a) searching for a moving target among still distractors is easier than searching for a still target among moving distractors; b) searching for a fast target among slow targets is easier than the opposite; c) adding variability in speed when searching for a unique motion direction has little effect, while adding variability in direction when searching for a unique speed makes the search task more difficult. The SSM is compelling because calculation of the saliency of objects is trivial and efficient, and because it has been shown to explain search results in experiments where dimensions other than motion are examined. We review this model in more detail in Section 3.

We enumerate the following challenges in transferring the existing perception knowledge to the problem addressed in this work:

- 1 In the perception experiments cited above, targets are chosen arbitrarily. In our experiment, targets are outliers in the statistical sense. We ask whether outlierness as a statistical property is preserved through the visual mapping.
- 2 Motion outlier detection in scatterplots is not a conjunction task. While the conjunction of motion and other dimensions is well studied, our problem is defined as a basic feature search in the presence of many irrelevant dimensions.
- 3 The dimensions in our stimuli encode continuous data attributes, while in perception studies they are often discretized to some degree (e.g., moving / still, fast / slow, bright / dim) [6, 21, 23, 33].

Animated Scatterplots

Scatterplots are one of the most effective visualizations because they employ position along a common scale, which was found to be the representation with which people can most accurately perform visual judgments [14]. Less important dimensions are commonly mapped to color, size, and shape. Gleicher et al. demonstrated that people can accurately compare means in multiclass scatterplots despite the addition of one discrete irrelevant cue (shape) [12]. This work shows that people can comfortably extract a summary statistic confined to a single dimension in the presence of an irrelevant dimension. Here, we investigate whether another summary statistic (outlierness) can be extracted from *motion* in correlated scatterplots with more than one irrelevant dimension (color, size). A key difference is that our scatterplots do not feature discrete dimensions that would enable the visual segmentation of the scene.

Szafir et al. argue that ensemble coding allows us to visually extract statistical information from scatterplots, such as outliers and statistical summaries, but acknowledge that attentional control may be problematic when multiple variables are encoded simultaneously, although the empirical

basis is still lacking [31]. Robertson et al. [24] found that animated scatterplots were not superior to static trend visualizations in analytical tasks (error rates) focused on trajectories. Huber and Healey [15] devised precise discriminability lower limits for motion (in displays with no competing visual channels): a target-distractor difference of at least 20 degrees is necessary for direction oddballs to be detected accurately; for speed, the difference needs to be at least 0.43 degrees of visual angle. Our outliers satisfy these conditions (Section 4).

Albeit designed to devise guidelines for notification design, Bartram et al.’s study of visual cues came to conclusions that relate to visualization design. Subjects were asked to perform a task in a window while glyphs overloaded with various encodings were scattered in the periphery [1]. The authors measured how accurately subjects could detect change in the glyphs. Motion was found to be the most reliable cue, better than changes in shape and color. They concluded that motion “does not seem to interfere with existing color and form coding” and that motion detection is effective even in visual periphery and with small amplitudes.

Etemadpour et al. [9, 10] used motion as a solution to clutter on the assumption that motion does not suffer interference from other channels. They reported a large improvement in the accuracy of ranking cluster density when motion was used as an encoding for cluster density. The improvements were relative to scatterplots where density was not explicitly encoded (implicitly encoded as position); plus, density is necessarily correlated to position, which makes motion-position a double encoding for density. Similarly, animated scatterplot matrices that encoded density with flickering were found superior to conventional ones in density judgement tasks [5].

3 SALIENCY

The statistical saliency model (SSM) [25] is a model of visual search based on the intuition that the visual system is interested in unusual things. Rosenholtz represents a visual scene in an appropriate feature space and then computes the saliency of a target as the number of standard deviations between its feature value and the mean of distractors. For a 1-D feature, this corresponds to a simple z-score, while for a higher number of dimensions, the saliency value is given by the Mahalanobis distance. Their model can be seen as a formalization of Duncan and Humphreys’ [8] rule of thumb that states that search is easier when target-distractor similarity decreases, or when distractor-distractor similarity increases.

The use of search tasks and reaction times as proxies for attention relies on the premise that search for salient items should be faster than search for items that do not draw attention. Rosenholtz’s study of visual search is directly relevant to motion outlier detection in visualization, and to ranking,

indirectly, if we assume that ranking points defaults to finding the most outlying point in increasingly narrow search spaces. For our purposes, however, the existing empirical validation of the SSM is limited. First, the scenes used to test it are usually distractor arrays of constant density (as in a uniform grid) [7]; second, no more than two features (speed and direction of motion) are varied. In information visualization displays, especially scatterplots, the x and y positions of points are commonly correlated, forming point clouds with varying density and levels of occlusion, and the points may be overloaded with multiple visual encodings, such as color, size, and shape [31].

A subsequent paper demonstrates how the SSM predicts asymmetries in colour search in the presence of non-neutral backgrounds [28]. The model is also the foundation for the feature congestion model of visual clutter [27], where separate pixel-level saliency maps of color and contrast luminance are linearly combined to produce clutter maps for raster images. The maps can be further aggregated to produce a scalar measure of overall display clutter.

Critically, it is not clear how low-level dimensions should be composed for the calculation of saliency in complex visualizations. In Rosenholtz’s study of motion outlier detection [25] it was suggested that the Mahalanobis distance should be calculated on the 2D space formed by speed and direction of motion, whereas in the feature congestion model saliency is calculated as a linear combination of 1D saliencies. It is likely that the latter is the appropriate method in a scene where motion and static features are varied, in which case we need to learn the dimension coefficients.

The pixel-level saliency maps employed in the feature congestion model and in many other saliency models [16] are not compelling for visualization applications because they operate after rendering, a late stage of the visualization pipeline, and because they are commonly tuned for natural images [3]. Recently, saliency models for data visualization were proposed [4, 20] that owe their performance mostly to accurate predictions of fixations on text elements (e.g., labels) in static visualizations.

In the next section, we will explain how we created stimuli with salient and non-salient targets following SSM’s definition of saliency.

4 EXPERIMENTAL DESIGN

We designed an experiment to find whether saliency predicts the accuracy of motion outlier detection tasks in animated multivariate scatterplots. In particular, we investigate whether saliency in irrelevant dimensions influences accuracy. Irrelevant dimensions are those that are not part of the task; for instance, when participants are instructed to find the fastest point, all dimensions (color, size, etc.) but speed are irrelevant.

The experiment is split into two *tasks*, a direction task and a speed task. The former asks participants to select the point with the most deviant direction, and the latter asks them to select the fastest point. Throughout this paper we will refer to visual channels as *dimensions*, and to specific values in these dimensions as *features*. We’ll also call direction and speed the *relevant dimensions* in their respective tasks. Each animated scatterplot (a *scene*) we produced has 12 *conditions*, where only the target is varied: a baseline where the target has no irrelevant salient features, plus five instances where it holds a single irrelevant salient feature (position, color, size, direction/speed, or size increase); a second baseline where the target has five irrelevant salient features at once, plus five instances where one irrelevant feature is held out. Thus, half the stimuli follows a one-at-a-time design, and the other half follows a hold-one-out design. We call these condition groups *saliency-deficient* and *saliency-charged*. We use the following notation to refer to individual conditions: in the saliency-deficient group, + conditions refer to the added irrelevant salient feature. For example, *+position* refers to a stimulus where the only irrelevant salient feature is position. In the saliency-charged group, - conditions refer to the removed irrelevant salient feature. For example, *-position* refers to a stimulus where only position is *not* salient. In all stimuli, the target has outlying value in the relevant dimension.





The reader may question why we do not vary dataset size, correlation, or the parameters of the sampling distribution. When distribution and dataset size are manipulated, the fundamental quantity that is being varied is the saliency of the target. For instance, a scene with more point spread results in less target saliency, and the same with a more crowded scene. As our goal is to find the effect of saliency on accuracy and we are already varying saliency by manipulating visual features, varying the factors in question would be redundant. Therefore, we see no reason in increasing the complexity of the experiment by adding these additional variables.

We generated ten different scenes per task, across 12 scene conditions, for a total of 120 stimuli per task. We collected 20 judgments of each stimulus for a total of 2400 judgments collected for each task, 200 per task-condition. We are interested in measuring the differences in error rates between the saliency-deficit and the saliency-charged baselines, and the impact of introducing or removing features.

Stimuli

We wrote a procedure for generating realistic stimuli inspired by animated scatterplots of the Gapminder data. The Gapminder plots map an often correlated pair of variables to the *x* and *y* coordinates, use size to encode a time-varying quantitative variable (usually population), and map a categorical

Table 1: Feature ranges. When speed is the task, the target is assigned an outlying distance value and mean or salient value for the other features. When direction is the task, the target receives an outlying direction. The color spectrum is defined by matplotlib’s Viridis colormap.

Feature	Range	Unit ≈	Mean ≈	Salient value	Outlier value
x, y	[0, 500]	px	250	variable	
color	[ , ]				
size (area)	[100, 600]	px	350	600	
size increase	[1, 2]	multiplier	1.5	2	
distance	[25, 100]	px	62.5	100	150
direction	[-81, 171]	degree	45	171	225

variable (continent) to color. In our scenes, we simulate instead a *continuous* variable mapped to color because it allows fine-grained control of the saliency.

A scene has 50 data points and is composed of two frames that are linearly interpolated to produce the animation. We decomposed motion into distance, which determines how much the point moves in the 2D plane (Euclidean distance), and direction. We sampled the features for the initial frame and calculated the positions in the final frame based on sampled values for distance and direction. x_1 and y_1 are sampled from a multivariate normal distribution with correlation 0.7. The values for color, size increase, and distance are sampled from independent normal distributions. Direction (angle) is sampled from a beta distribution ($\alpha = 9.55, \beta = 10$) that has shape similar to a normal, but produces values that are more concentrated around the mean. This pattern was chosen to preserve the correlation of the plot; that is, the point cloud, *as a whole*, should be moving in a well-defined direction. Due to the animation duration being constant for all points, distance is effectively a measure of speed.

After all points are sampled, a target is selected according to the condition. If position is salient, then we select the point with the highest Mahalanobis distance (i.e., the most distant from the center of the point cloud); otherwise, the point *closest* to the center is selected. If color is salient, we assign to the target the maximum color in the color range; otherwise, we assign it the mean color. This pattern is followed for all the other irrelevant visual dimensions.

All targets are outliers detectable through the interquartile range method (Tukey’s fences, $k=1.5$); thus, an analyst using boxplots to analyze the distributions of speed and direction would clearly identify the target as an outlier (positioned beyond a boxplot’s whiskers). We produced outliers by assigning to targets a constant value outside the sampled distribution range. On average, direction and speed outlier

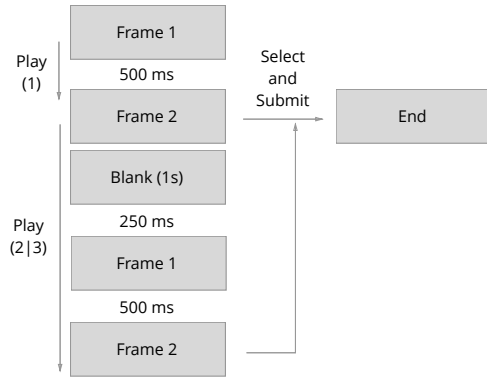


Figure 3: Flow diagram illustrating the sequence of screens in the study interface. Participants could replay the animations twice. Blank screens were placed in-between replays.

values were 3.11 and 3.82 standard deviations from the mean. For comparison with Huber and Healey’s discriminability thresholds, in average, the trajectory of speed outliers was 0.95 degrees of subtended visual angle longer (40% higher) than that of the next fastest point on a 113ppi laptop screen (e.g., Macbook Pro 13in.) at typing distance (20in.). The difference between direction outliers and the next most deviant points was 52 arc degrees (43% higher), in average.

Table 1 lists the dimension ranges for the sampled points, as well as the mean and salient values. We use the inverted version of matplotlib’s Viridis colormap [30], where higher values are darker (bright points on a white background would not “pop out”). Viridis was found to have superior performance, measured in time and accuracy of relative similarity judgments, in comparison with other popular colormaps [19]. We chose the direction range again respecting the principle that the plot trend shouldn’t be overly disrupted. The size range was chosen so as not to cause too much occlusion. In addition, the render order on the screen (from largest to smallest) also reduced occlusion. We inspected the stimuli to make sure that the targets were not occluded. Size increase is a multiplier of the initial size. Figure 1 displays a scene for the direction task in the saliency-deficient baseline condition. The target moves in an outlying direction but has average values for speed, color, size and position. All stimuli are provided in the supplemental materials.

Procedure

We presented the stimuli embedded in the Mechanical Turk interface (Figure 2). The page presented the first frame of the animation until the play button was pressed. After the end of the animation, the visualization was stationed in the second frame, allowing participants to select the target and submit the response or replay the animation up to two times before submission. The animation duration was 500 milliseconds.

When play was pressed the second or third time, the points faded to a blank screen then reappeared in their first frame positions before the animation took place. This sequence is illustrated in Figure 3. The variable number of views was introduced as a measure to mitigate errors due to interruptions, as these can be a problem in crowdsourced studies where we have no control over the environment. The number of views was capped at three to prevent the task from becoming too easy to the extent no differences can be detected between the conditions. Trials were published as two separate groups of HITs on Mechanical Turk (speed and direction). Within each group, trials appeared in random order. Participants were not limited in the number of tasks they could complete. We recorded time, accuracy and number of views.

Participants were instructed to find the fastest point (“find the fastest point”) in the speed task and the most deviant point (“find the point that has the most unique trajectory compared to the rest”) in the direction task. Therefore, the task is to “find the maximum”, with all targets being outliers. This mitigates the risk of participants not comprehending the outlieriness concept or the study being affected by different notions of what an outlier is. Participants had the opportunity to perform test trials, as it is common on MTurk, but these trials did not provide feedback.

5 EXPERIMENT RESULTS

We collected 4800 observations from 67 participants, who performed an average of 71.6 tasks ($sd = 42.4$). The median completion time was 10.3s. Figure 4 displays the accuracy distribution per task-condition. Accuracy is calculated per stimulus (a scene-condition pair) as the ratio correct/incorrect. In the following sections, we examine the odds of a participant selecting the outlier and which features contributed most to incorrect selections.

Channel Contributions

We used the R package lme4 [2] to fit a pair of generalized linear mixed models (GLMM), one for each task (speed and direction). We specified the models with a binary response variable (correct = [true, false]) and five binary covariates [salient, non-salient]: position, color, size, speed/direction, and size increase. This model is also known as a binomial logistic regression. In order to account for scene-specific and participant-specific effects, we inserted the variables *scene* and *subject* as random effects. As such, the random impact from scenes that happen to be more or less difficult, or participants that are more or less accurate, is reduced. Figure 4 shows the data, and Figure 5 shows the model estimates. Below we discuss the main findings.

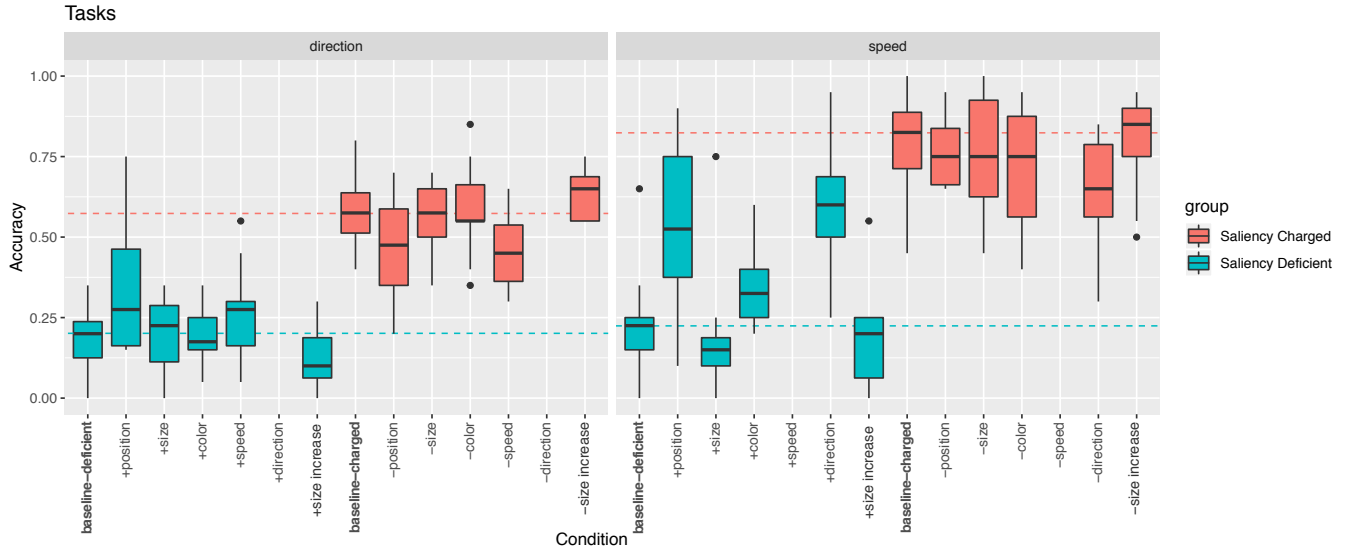


Figure 4: Distribution of accuracy for each condition. For each stimulus, accuracy is calculated over 20 judgments. There are 10 stimuli per condition, one for each scene.

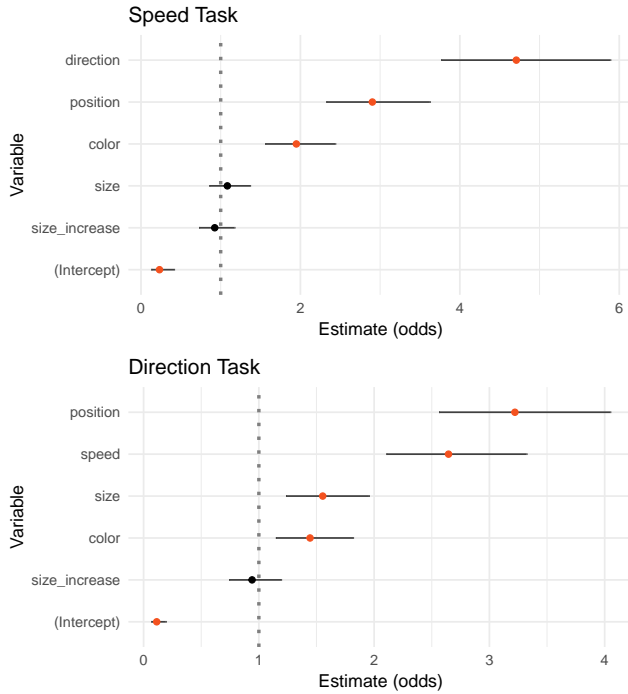


Figure 5: Estimates for the effect of irrelevant salient features on the odds of a speed (top) and direction (bottom) outlier being identified. Binary covariates and multiplicative coefficients. Red denotes statistical significance ($p < .05$).

Motion outlier detection is not well supported. We observed a mean accuracy lower than 25% in the condition *baseline-deficient* in both tasks. This condition is where the motion

outlier does not have salient features other than motion. Low accuracy suggests subjects were mostly unable to separate motion from other dimensions in order to correctly identify the motion outlier. In other words, motion detection in multivariate scatterplots suffers interference from irrelevant dimensions.

Accuracy depends on the saliency of irrelevant features. We recorded higher accuracy in most conditions where the motion outlier had irrelevant salient features. In particular, subjects achieved averages of 78.5% and 58.5% accuracy in the *baseline-charged* condition, in the speed and direction tasks, respectively. Removing one salient feature at a time generally caused a drop in accuracy; conversely, adding one salient feature generally increased accuracy, but not by much, especially in the direction task, which suggests that in crowded displays motion outliers can only reliably be extracted if they have multiple salient features. More generally, animated scatterplots may reliably support only the detection of *global* outliers.

Direction plays the largest role in the speed task. The fitted model indicates that direction saliency accounts for an increase of 4.7 times in the odds of correct speed outlier detection, which corresponds to a shift in probability from 0.19 (intercept) to 0.52. This result is somewhat aligned with previous findings that direction variability degrades searching for a unique speed. Targets with salient direction might have allowed subjects to segment the scene, cancelling some of the noise that impacts accuracy.

Position plays the largest role in the direction task accuracy. Position is estimated to account for an increase of 3.2 times in the odds for the direction task, which is equivalent to a shift in probability from 0.10 (intercept) to 0.26. This result is not trivial: while targets in salient positions (surrounded with blank space) are more visible, they are arguably more difficult to compare, due to their distance from other points. In addition, this result highlights the effect of clutter on this task. Our sampling process produces a point cloud with a high-density center. Points with low spatial saliency are located in these cluttered regions.

Size and color have small influence in the direction task. Both size and color contributed modestly to the outcome. We found no evidence of a difference between the odds estimate for these dimensions, as their confidence intervals largely overlap. In general, we observe precedence of spatial attributes (position, speed, and direction) over form attributes (color and size).

Size makes no difference in the speed task. We found that size and size increase did not alter the odds of correct detection in the speed task (these variables have odds ratio approximately 1). This is in contrast to a small, but significant effect in the direction task. It is possible that this can be explained by larger points being perceived as moving slower, which would degrade the performance relative to the baseline; however, our model did not point to a negative effect. It is also plausible that the distribution of values mapped to size did not produce enough saliency. Weber's law predicts a non-linear relation between area change and perceived area change, which may have caused points with maximum area to appear closer to the mean and less salient.

Which features mislead?

When examining the incorrect choices of participants one would normally expect that the points they selected are close to the target in speed or direction; that is, more incorrect selections should be recorded for faster or more deviant points. This expectation was contradicted by the low correlations we observed between task dimensions and selection counts: 0.23 for speed and 0.28 for direction. The correlations were calculated on the subset of non-target points with selection count greater than 0. This suggests that incorrect selections are not necessarily due to the proximity to the outlier value in the target dimension; that is, irrelevant dimensions may be leading participants to make mistakes.

To find which dimensions play a role in the number of times a non-target point is selected we fit generalized linear models (GLM) to the subset of 1,530 non-target points that were selected at least once. Since the observed response variable—selection count—is skewed and lies in the interval $(1, \infty)$ we set the models with a Gamma response variable.

The covariates are saliency measures (SSM) on speed/direction and on all other dimensions. We use the saliency measure here because unlike targets, which were made either salient or not, non-target features lie within a saliency spectrum. Likewise, we split position saliency into saliency in the first frame (xy_1) and in the second frame (xy_2).

We included terms for interactions of all saliency measures with speed/direction. In order to make the estimates comparable and easier to interpret we standardized all covariates (zero-mean and unit-variance). In Figure 6, the effects are multiplicative; that is, $y = \beta_0 \times \beta_1 x_1 \times \beta_2 x_2 \times \beta_{12} x_1 x_2 \dots$, where β_0 is the intercept, β_i are fixed effects, β_{ij} is an interaction term, and x_i are dimension values. The interaction plots in Figure 6 depict the curve that represents the relationship between speed/direction and the response variable (count), and how this curve is changed as a function of the interacting variable. Below we report the main findings.

Position and direction saliencies boost the effect of speed. In the speed task, the model estimates reveal, not surprisingly, that speed is a confuser and that the interactions of speed with direction saliency and position saliency in the first frame are significant. The interaction terms are positive: the misleading effect of speed increases as a function of the saliency of these irrelevant dimensions. In Figure 6 (top right), this is shown as an increase in slope: when the values of either direction saliency or position saliency increase by one standard deviation, the effect of speed on the response becomes steeper. In practice, this indicates that fast points moving from blank regions and in unique directions tend to be mistaken for true speed outliers. This result is aligned with the channel contributions observed in the previous section: position and direction have the highest impact on the odds of a target being correctly identified.

Position saliency in the first frame and color saliency boost the effect of direction. In the direction task we found that the misleading effect of direction saliency is boosted by position saliency in the first frame. In Figure 6 (bottom right) this is seen as a slope increase when the value of *saliency_xy1* increases. Color saliency also interacts with direction, but to a lesser extent. In addition, we found that the effect of speed is significant and independent from that of direction. Considering the results above, it appears that position saliency in the first frame is consistently a major factor for selection. Motion outliers that are inside the point cloud might be overlooked if there is a confuser departing from a salient position.

Position saliency in the second frame degrades the effect of direction. Surprisingly, we found that position saliency in the second frame has a negative interaction with direction. This appears in Figure 6 as a decrease in the slope of the curve when *saliency_xy2* increases. Participants are thus less likely

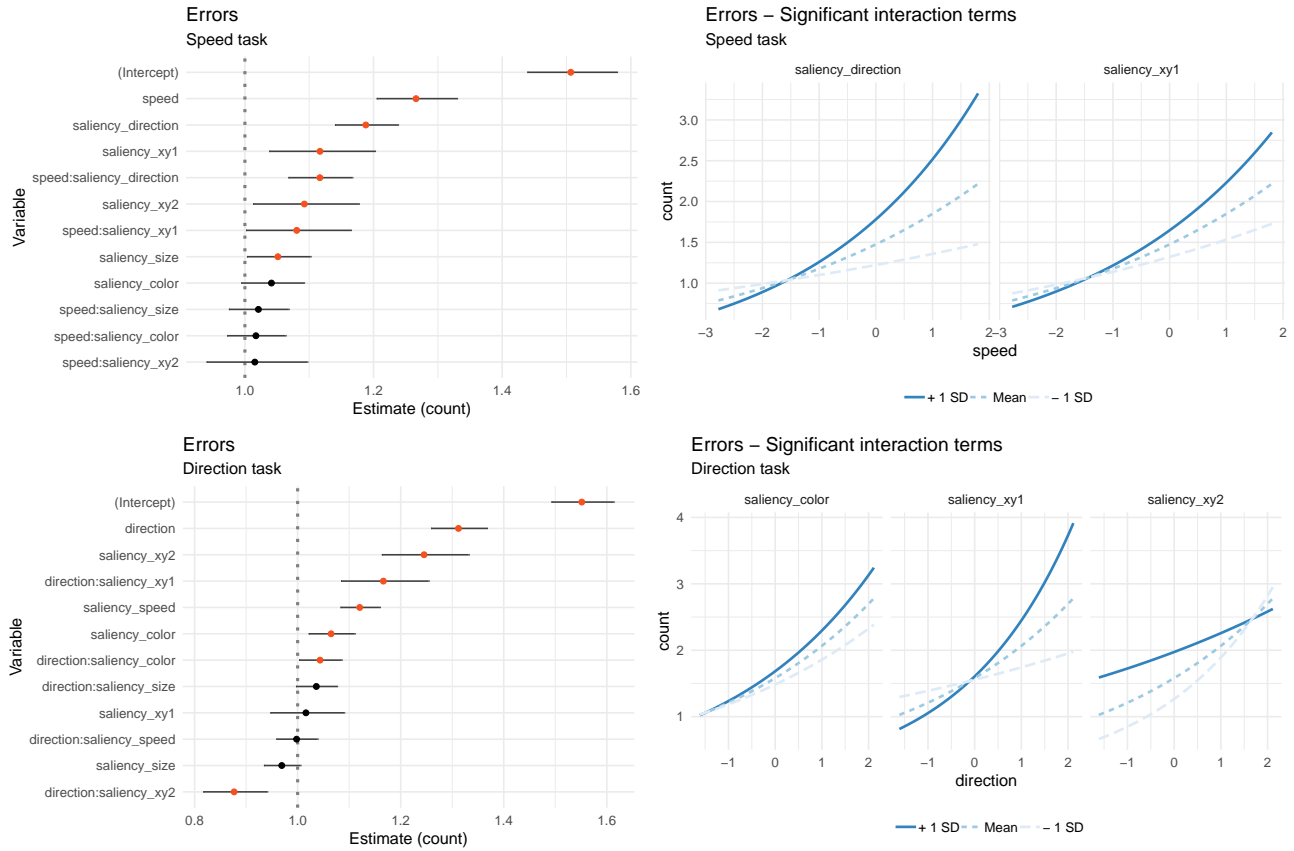


Figure 6: Left: estimates for the effect of feature saliency on the number of times a non-target is selected (erroneously). Right: interaction plot depicting the modulation of the effect of speed and direction by irrelevant features.

to erroneously select a point moving in a salient direction the more salient its final position. We hypothesize that this effect may be due to points moving out of the cloud clearly having direction perpendicular to the trend. As participants were instructed to select “the point that moves in the most deviant direction”, they may have been looking for points that were in the opposite direction of the mean. Points moving in the opposite direction would likely be inside the cloud, not moving out of it.

Replays

In this section, we examine the number of times participants viewed the animation before selecting their answers. We analyze the distribution of correct and incorrect selections across the three possible values for number of views. Figure 7 shows this distribution split by task, condition, and whether the trial was completed correctly. Due to the study being deployed on Mechanical Turk, we are unable to separate divided attention from task difficulty as the cause for replays. A reproduction of this experiment in a controlled setting is necessary for establishing a causal relationship.

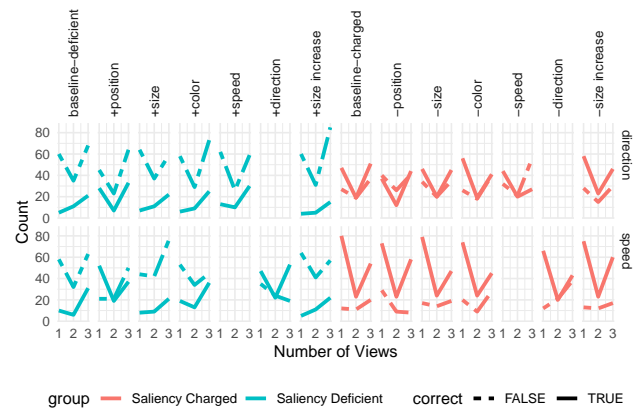


Figure 7: Distribution of number of views divided by task and condition.

Overall, we observe the prevalence of a V-shaped distribution, suggesting that participants were more likely to watch the animation either the minimum or maximum allowed times. In the saliency charged group, speed task, we see a clear pattern of correct answers coming more often from 1-view judgments. This pattern is not present in the direction

task. In the saliency deficient group we observe the opposite pattern: correct judgments are more likely to come from 3-view judgments, with a few exceptions; namely, targets with salient position in the direction and speed tasks and with salient direction in the speed task seem to require less effort than targets in the other deficient conditions. These patterns are consistent with the coefficients found in the above analyses, suggesting task difficulty may be behind them. We also see the V-shaped symmetrical pattern in incorrect answers, especially in the direction tasks, suggesting confidence in wrong selections.

6 DISCUSSION

We found that motion outlier detection is unreliable in multivariate animated scatterplots. The accuracy of motion outlier detection is degraded in the absence of other salient cues. This suggests a level of interference of spatial (position) and form (color, size) encodings over motion, and between the individual components of motion (speed and direction). Furthermore, we found evidence that while people were selecting outliers based on the relevant features (speed, direction), irrelevant features may have acted as “boosters,” leading people to select the wrong target. We hypothesize that this may be due to people’s attention getting caught by near-outliers that have high global saliency; since the animation is short, they would not have enough time to revise a first impression.

We found spatial saliency, which is closely tied to clutter, to have a large impact on accuracy in both speed and direction tasks. Here, we emphasize the distinction between occlusion and clutter. We inspected the stimuli for occlusion and adjusted the z-order of elements to prevent small points hiding under larger points. Instead of an effect due to inability to *see* the targets, we believe the effect is due to a difficulty of *allocating attention*, in the sense of feature congestion: as the feature space becomes crowded there is less chance for a single object to stand out [26].

The results suggest that it may be possible to predict scenes where outlier detection is difficult on the basis of saliency measurements. A linear model with a binary response variable and feature saliency coefficients such as the one we fit can output the odds of correct detection given a “scene.” A linear model of saliency (for clutter measurement) was used also by Rosenholtz et al. [26]. A threat to the generality of this approach is the fact that the statistical saliency model is invariant to scale (due to the use of Mahalanobis distance); for instance, points mapped to a very narrow color range yield the same saliency values as if they were mapped to a wide color range.

At a more general level, the results expose a failure of mapping data outliers to visual outliers, which we refer to as a *saliency deficit*. A data point or a group of data points is saliency deficient when its importance in the data space

is not reflected in the visualization due to a lack of saliency. Saliency deficit is thus a condition of imbalance between data and visual importance. In Kindlmann and Scheidegger’s algebraic model [18], such a failure is classified as a violation of the visual-data correspondence principle: important changes in the data should yield important visual changes.

The notion of saliency deficit is task dependent: here we examined motion outlier detection, but it is possible that other tasks in other visualization types may suffer from the same problem. Interference between visual channels is not new in visualization research, which often points to the theory of separable and integral dimensions [11]. When a pair of visual dimensions is integral, information from an individual dimension cannot be accessed easily. However, these studies have been traditionally restricted to the task of class-separation and with static features. For instance, in a point cloud with varying hue and size, it’s not easy to separate points based on each dimension independently.

It is plausible that the mechanism behind saliency deficit depends on the number of visual channels employed. That is, the more visual channels, the harder it becomes to perform tasks that rely on saliency along a single dimension. This sends us back to the feature congestion model of clutter, which predicts difficulty in creating salient targets within a crowded feature space. In order to assert this mechanism with confidence, further research needs to examine this effect with a variable number of visual channels.

7 LIMITATIONS

We would like to see the present experiment extended in many ways. We controlled the outlierness of the targets, animation speed, and the distribution of the features and their correlation in order to isolate the effect of feature saliency. This imposes limitations on the scope of inference of the experiment. It is plausible that interactions exist between the controlled factors and the response variables; in particular, as the outlierness of the target increases, the effect of other features probably decreases. The effect of animation speed may be complex: fast transitions may make tasks more difficult, but studies in the topic of change blindness have found that large changes can also go undetected when introduced gradually [29].

We have investigated only positive outliers. Due to a known asymmetry in motion target detection—it is easier to find fast targets among slow distractors than the inverse—we cannot extend our conclusions to slow outliers.

As stated in Discussion, we would like to measure accuracy in an experiment where the number of irrelevant dimensions is manipulated. This could generate insights on the number of dimensions beyond which some tasks start to lose accuracy. Likewise, it would be interesting to measure the effect of motion on other encodings. Finally, it is possible that the

estimates for size and color do not generalize to other ranges. In particular, the color saliency may vary depending on the direction of the colormap (bright to dark or inverse) and the background.

8 CONCLUSION

We reported the results of a controlled experiment designed to test the effect of irrelevant visual dimensions on the accuracy of motion outlier detection in multivariate animated scatterplots. We found that color, size, position, speed, and direction influence the accuracy with which people detect the fastest or the most deviant data point. In particular, we found that spatial visual dimensions, such as position, speed, and direction have larger influence than form attributes, such as color and size. Mean accuracy in detection of speed outliers was higher than 75% only when targets had multiple salient features. When detecting direction outliers, mean accuracy was never higher than 30% when targets lacked salient features.

These results suggest a saliency deficit effect that prevents motion targets from being detected accurately when their overall saliency is low; as a consequence, animated scatterplots should be used with caution if outlier detection is a critical task. We believe saliency deficit may affect tasks in other multivariate visualizations. Models of task accuracy that rely on foundational variables, such as saliency, in conjunction with models of user intent may inform the introduction of automated interventions when the predicted accuracy of a task given a plot is low.

ACKNOWLEDGEMENTS

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) and Fundação CAPES (9078-13-4/Ciência sem Fronteiras).

REFERENCES

- [1] Lyn Bartram, Colin Ware, and Tom Calvert. 2003. Moticons:: detection, distraction and task. *Int. Journal of Human-Computer Studies* 58, 5 (2003), 515–545.
- [2] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48.
- [3] Zoya Bylinskii, Tilke Judd, Ali Borji, Laurent Itti, Frédo Durand, Aude Oliva, and Antonio Torralba. [n. d.]. MIT Saliency Benchmark. <http://saliency.mit.edu/>.
- [4] Zoya Bylinskii, Nam Wook Kim, Peter O'Donovan, Sami Alsheikh, Spandan Madan, Hanspeter Pfister, Fredo Durand, Bryan Russell, and Aaron Hertzmann. 2017. Learning visual importance for graphic designs and data visualizations. In *Proc. 30th ACM Symp. on User Interface Software and Technology*. ACM, 57–69.
- [5] Helen Chen, Sophie Engle, Alark Joshi, Eric D Ragan, Beste F Yuksel, and Lane Harrison. 2018. Using Animation to Alleviate Overdraw in Multiclass Scatterplot Matrices. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 417.
- [6] Lisa J Croner and Thomas D Albright. 1997. Image segmentation enhances discrimination of motion in visual noise. *Vision Research* 37, 11 (1997), 1415–1427.
- [7] M Dick, S Ullman, and D Sagi. 1987. Parallel and serial processes in motion detection. *Science* 237, 4813 (July 1987), 400–402.
- [8] J. Duncan and G. W. Humphreys. 1989. Visual search and stimulus similarity. *Psychological Review* 96, 3 (1989), 433–458.
- [9] Ronak Etemadpour and Angus Graeme Forbes. 2017. Density-based motion. *Information Visualization* 16, 1 (2017), 3–20.
- [10] Ronak Etemadpour, Paul Murray, and Angus Graeme Forbes. 2014. Evaluating density-based motion for big data visual analytics. In *Proc. of the IEEE Int. Conf. on Big Data*. IEEE, 451–460.
- [11] Wendell R Garner. 2014. *The processing of Information and Structure*. Psychology Press.
- [12] M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri. 2013. Perception of average value in multiclass scatterplots. *IEEE Trans. on Visualization and Computer Graphics* 19, 12 (Dec. 2013), 2316–2325.
- [13] Christopher Healey and James Enns. 2011. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Trans. on Visualization and Computer Graphics* (July 2011), 1–20.
- [14] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using Mechanical Turk to assess visualization design. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems*. 203–212.
- [15] Daniel E Huber and Christopher G Healey. 2005. Visualizing data with motion. In *Visualization, 2005. VIS 05. IEEE*. IEEE, 527–534.
- [16] Tilke Judd, Frédo Durand, and Antonio Torralba. 2012. *A Benchmark of Computational Models of Saliency to Predict Human Fixations*. Technical Report MIT-CSAIL-TR-2012-001.
- [17] Younghoon Kim and Jeffrey Heer. 2018. Assessing effects of task and data distribution on the effectiveness of visual encodings. *Computer Graphics Forum (Proc. EuroVis)* (2018).
- [18] Gordon Kindlmann and Carlos Scheidegger. 2014. An algebraic process for visualization design. *IEEE Trans. on Visualization and Computer Graphics* 20, 12 (2014), 2181–2190.
- [19] Yang Liu and Jeffrey Heer. 2018. Somewhere over the rainbow: An empirical assessment of quantitative colormaps. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems*.
- [20] Laura E Matzen, Michael J Haass, Kristin M Divis, Zhiyuan Wang, and Andrew T Wilson. 2018. Data Visualization Saliency Model: A Tool for Evaluating Abstract Data Visualizations. *IEEE Transactions on Visualization & Computer Graphics* 1 (2018), 563–573.
- [21] Peter McLeod, Jon Driver, and Jennie Crisp. 1988. Visual search for a conjunction of movement and form is parallel. *Nature* 332, 6160 (1988), 154.
- [22] D. Moritz, C. Wang, G. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. 2019. Formalizing visualization design knowledge as constraints: actionable and extensible models in Draco. *IEEE Trans. on Visualization and Computer Graphics* (2019).
- [23] Thomas V Pappathomas, Andrei Gorea, and Bela Julesz. 1991. Two carriers for motion perception: color and luminance. *Vision Research* 31, 11 (1991), 1883–1892.
- [24] George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, and John Stasko. 2008. Effectiveness of animation in trend visualization. *IEEE Trans. on Visualization and Computer Graphics* 14, 6 (2008), 1325–1332.
- [25] Ruth Rosenholtz. 1999. A simple saliency model predicts a number of motion popout phenomena. *Vision Research* 39 (1999), 3157–3163.
- [26] Ruth Rosenholtz, Yuanzhen Li, Jonathan Mansfield, and Zhenlan Jin. 2005. Feature Congestion: A Measure of Display Clutter. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. 761–770.
- [27] Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. 2007. Measuring visual clutter. *Journal of Vision* 7, 2 (2007), 17.1–22.

- [28] Ruth Rosenholtz, Allen L. Nagy, and Nicole R. Bell. 2004. The effect of background color on asymmetries in color search. *Journal of Vision* 4, 3 (March 2004), 224–240.
- [29] Daniel J. Simons, Steven L. Franconeri, and Rebecca L. Reimer. 2000. Change blindness in the absence of a visual disruption. *Perception* 29, 10 (2000), 1143–1154.
- [30] Nathaniel Smith and Stéfan van der Walt. 2015. A better default colormap for Matplotlib. <https://youtu.be/xAoljeRJ3lU>
- [31] Danielle Albers Szafr, Steve Haroz, Michael Gleicher, and Steven Franconeri. 2016. Four types of ensemble coding in data visualizations. *Journal of Vision* 16, 5 (2016), 11.
- [32] Anne M Treisman and Garry Gelade. 1980. A feature-integration theory of attention. *Cognitive Psychology* 12, 1 (1980), 97–136.
- [33] Adrian Von Mühlenen and Hermann J Müller. 2000. Perceptual integration of motion and form information: evidence of parallel-continuous processing. *Perception & Psychophysics* 62, 3 (2000), 517–531.
- [34] Colin Ware. 2004. *Information Visualization: Perception for Design* (2nd ed.). Morgan Kaufmann Publishers Inc., San Francisco.
- [35] Jeremy M Wolfe. 1998. What can 1 million trials tell us about visual search? *Psychological Science* 9, 1 (1998), 33–39.